AD-A079 247 ARMY PERSONNEL RESEARCH OFFICE WASHINGTON DC F/G 5/9

A TECHNIQUE FOR SIMULATING MIXED NORMAL AND UNIFORM DISTRIBUTIO--ETC(U)

DEC 65 R F BOLDT

UNCLASSIFIED APRO-RM-65-9 NL

| OF |

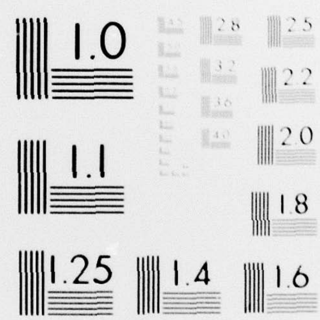
AD A079247

END

DATE FILMED

2-80

DDC




MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

ADA079247

28 February 1966

MASTER FILE
Office, Chief [illegible]
U.S. [illegible], OCRD

LEVEL II

1

9 Research Memorandum 65-9

6

# A TECHNIQUE FOR SIMULATING MIXED NORMAL AND UNIFORM DISTRIBUTIONS

14 APRO-RM-65-9

11 December 1965

12 10

10 Robert F. Boldt

DDC RECEIVED DEC 19 1979 A

DISTRIBUTION STATEMENT A
Approved for public release;
Distribution Unlimited

DDC FILE COPY

79 12 18 201

U. S. ARMY PERSONNEL RESEARCH OFFICE

040 650

Army Project Number
2J024701A723

MICS b-11

Research Memorandum 65-9

# A TECHNIQUE FOR SIMULATING MIXED NORMAL AND UNIFORM DISTRIBUTIONS

by Robert F. Boldt

Richard C. Sorenson, Task Leader

Submitted by:
Cecil D. Johnson
Chief, Statistical Research
and Analysis Laboratory

Approved by:
J. E. Uhlaner
Director of
Laboratories



December 1965

Research Memorandums are informal reports on technical research problems. Limited distribution is made, primarily to personnel engaged in research for the U. S. Army Personnel Research Office.

## A TECHNIQUE FOR SIMULATING MIXED NORMAL AND UNIFORM DISTRIBUTIONS

Because of the multivariate nature of personnel problems, simulation of manpower systems requires techniques for generating sets of variables with prespecified joint distribution functions. For example, Sorenson[1], while studying the impact of manpower allocation policies, has used the joint normal distribution to simulate the enlisted input. Techniques used to generate these scores have been described by Boldt, Wiskoff, and Fitch[2] and by Hillier[3]. For many problems, the normal approximation will be quite adequate.

For certain problems, however, a procedure which allows the generating of both uniform and normal variables is desirable. For example, when the Armed Forces Qualification Test (AFQT) score is included with Army Classification Battery (ACB) scores, the uniform distribution is introduced since the AFQT, being a percentile measure, is uniformly distributed. Conceivably, other percentile measures could be introduced also. Hence the problem of generating mixed normal and uniform distributions can have application in personnel simulations. Procedures for generating such distributions are presented below and their derivations outlined.

### CASE OF A SINGLE UNIFORM VARIABLE

Since procedures are available for generating joint normal distributions with specified means and variance-covariance matrices, it will be useful to apply them. A brief review of these procedures follows. Means will not be considered in this paper, since they can be easily incorporated with an additive adjustment.

Suppose it is desired to generate a set of variables jointly normally distributed with variance-covariance matrix S. This is accomplished by generating a set of variables which are jointly normally distributed with variance-covariance matrix equal to the identity matrix to get the matrix X. Then find T such that $T'T = S$. Then the score matrix XT will contain variables which are jointly normally distributed with variance-covariance matrix S. This procedure depends on the fact that linear sums of jointly normally distributed variables are themselves jointly normally distributed.

---

[1] Sorenson, R. C. Optimal allocation of enlisted men--full regression equations vs. aptitude area scores. U. S. APRO Technical Research Note 163. November 1965.

[2] Boldt, R. F., Wiskoff, M. F., Fitch, D. J. An allocation technique applied to current aptitude input. U. S. APRO Research Memorandum 60-19. November 1960.

[3] Hillier, A. A program for computing probabilities over regions under the multivariate normal distribution. Tech. Rep. No. 54, Applied Mathematics and Statistics Laboratory, Stanford University, 1961.

The procedure suggests the possibility that to mix normals and uniforms, with prescribed variance-covariance matrices, the normals and uniforms could be generated independently and then transformed linearly to get the prescribed variance-covariance matrices. Such a procedure would indeed produce variables with the proper variance-covariance matrices, but the marginal distributions would be neither normal nor uniform since linear composites of mixed distributions do not, in general, retain the original distributions. Hence, such a procedure is not acceptable.

The method suggested here is to generate normal variables and transform them to a distribution whose variance-covariance matrix is consistent with the variance-covariance matrix desired when certain of the variables are then transformed back into percentile measures. Consider, for example, two variables which are jointly normally distributed with correlation $\rho$. Suppose one of these variables is then transformed into a uniform distribution to get a variable which correlates r with the other normally distributed variable. There is clearly a functional relationship between $\rho$ and r. If the value of r were known empirically, then one could find $\rho$, generate two normally distributed variables with correlation $\rho$, transform one to the uniform distribution to get two variables, one normal and one uniform, with the desired correlation r. The functional relation, whose derivation is outlined in Appendix A, has been found and is

$$r = \rho \sqrt{\frac{3}{\pi}} \cong \rho\ .9772. \qquad (1)$$

Hence, as one might expect, the transformation from normal to uniform has negligible effect on the correlation. However, knowing the column vector of correlations of the normally distributed variables with the uniformly distributed variable, we compute $(1/.9772)$ times these correlations to get the desired correlations for normally distributed variables. We then generate all jointly normally distributed variables and transform the variable whose correlations have been adjusted into a uniform distribution. For example, this procedure is suggested when an AFQT type variable is used with ACB type variables.

## CASE OF MORE THAN ONE UNIFORM VARIABLE

For many uniform variables, an analogous procedure is suggested. In this case, though, one also needs to find the relationship between the correlation coefficients of two percentile variables, R say, which would be obtained if two normal variables with correlation $\tilde{R}$ were transformed into uniformly distributed variables. This relationship, whose derivation is outlined in Appendix B, is given below.

$$R = (6/\pi)\ \text{Arc Tan}\ \left( \tilde{R} / \sqrt{4 - \tilde{R}^2} \right) \qquad (2)$$

Hence, for each value of R, a value of $\tilde{R}$ can be found which can be used in the matrix S.

## DISCUSSION

Examination of equation (1) indicates that the correlation coefficients are not greatly affected by transformation from normal to uniform. Analysis of equation (2) indicates that the maximum difference between $\tilde{R}$ and R is approximately .027. Hence, one can place good reliance on studies in which the adjustments were not made. However, to add an extra bit of representational validity to future simulation studies, it is suggested that the corrections be applied.

## APPENDIX A

Functional Relationship between r and $\rho$

Let x and y be jointly normally distributed with means zero and variances equal to one. Then

$$q = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} \exp\left[-z^2/2\right] dz$$

is uniformly distributed. The expected correlation of y with q, r, is given by

$$r = \frac{E[(q - .5)y]}{1/\sqrt{12}} \tag{3}$$

where E is the usual expectation operator.

The figures .5 and $1/\sqrt{12}$ are the mean and standard deviation, respectively, of q. Since the mean of y is zero,

$$r = \sqrt{12}\, E(qy).$$

By definition,

$$E(qy) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \frac{y}{2\pi\sqrt{1-\rho^2}} \left[\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} \exp\left[-z^2/2\right] dz\right] \exp\left[-(x^2-2\rho xy+y^2)/2(1-\rho^2)\right] dx\, dy. \tag{4}$$

The following sequence of transformations is applied.

$$w = z - x \qquad dw = dz$$

$$u = \frac{y - \rho x}{\sqrt{1-\rho^2}} \qquad du\sqrt{1-\rho^2} = dy$$

Then

$$E(qy) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{0} \frac{\rho x + u\sqrt{1-\rho^2}}{(2\pi)^{3/2}} \exp\left[-w^2 + 2xw + 2x^2 + u^2)/2\right] dw\, dx\, du$$

$$= \int_{-\infty}^{\infty}\int_{-\infty}^{0} \frac{\rho x}{2\pi} \exp\left[-(w^2 + 2xw + 2x^2)/2\right] dw\, dx$$

Letting

$$v = x\sqrt{2} + w/\sqrt{2} \qquad dx = \frac{1}{\sqrt{2}}\, dv$$

and integrating out x yields

$$E(qy) = -\int_{-\infty}^{0} \frac{\rho w}{2\sqrt{2\pi}} \exp\left[-w^2/4\right] dw$$

Let $\sqrt{2}\, m = w$ and $dw = \sqrt{2}\, dm$

yields

$$E(qy) = -\int_{-\infty}^{0} \frac{\rho m}{2\sqrt{\pi}} \exp\left[-m^2/2\right] dm = \rho \frac{1}{2\sqrt{\pi}}$$

and $$r = \sqrt{12}\, E(qy) = \rho\sqrt{\frac{3}{\pi}}\,.$$

## APPENDIX B

### Functional Relationship between R and $\widetilde{R}$

Let x and y be jointly normally distributed. Then

$$q = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{z^2}{2}} dz \quad \text{and} \quad p = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{y} e^{-\frac{z^2}{2}} dz$$

are uniformly distributed. Their correlation is

$$\begin{aligned} R &= [E(p - .5)(q - .5)]12 \\ &= [E(pq) - .25]12. \end{aligned} \tag{5}$$

By definition

$$E(pq) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \frac{1}{2\pi\sqrt{1 - \widetilde{R}^2}} \left[ \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{z^2}{2}} dz \right] \left[ \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{y} e^{-\frac{z^2}{2}} dz \right] e^{-\frac{(x^2 - 2\widetilde{R}xy + y^2)}{2(1-\widetilde{R}^2)}}$$

Applying the following sequence of transformations,

$$\begin{aligned}
u &= z - x & du &= dz \\
v &= z - y & dv &= dz \\
x &= y + 2\beta & dx &= 2d\beta \\
y &= \alpha - \beta & dy &= d\alpha
\end{aligned}$$

$$\alpha \sqrt{\frac{2(2+\widetilde{R})}{1+\widetilde{R}}} + (u+v) \sqrt{\frac{1 + \widetilde{R}}{2(2+\widetilde{R})}} = W \qquad d\alpha = \sqrt{\frac{1 + \widetilde{R}}{2(2+\widetilde{R})}}\, dW$$

$$\beta \sqrt{\frac{2(2-\widetilde{R})}{1 - \widetilde{R}}} + (u+v) \sqrt{\frac{1-\widetilde{R}}{2(2-\widetilde{R})}} = \gamma \qquad d\beta = \sqrt{\frac{1 - \widetilde{R}}{2(2-\widetilde{R})}}\, d\gamma$$

$$a = u\sqrt{\frac{2}{4 - \widetilde{R}^2}} \qquad du = da\sqrt{\frac{4 - \widetilde{R}^2}{2}}$$

$$b = v\sqrt{\frac{2}{4 - \widetilde{R}^2}} \qquad dv = db\sqrt{\frac{4 - \widetilde{R}^2}{2}}$$

and integrate out w and $\gamma$ to get

$$E(pq) = \int_{-\infty}^{0}\int_{-\infty}^{0} \frac{\sqrt{4 - \widetilde{R}^2}}{2^2\,\pi}\; e^{-\frac{1}{2}\left(a^2+b^2-2ab\frac{\widetilde{R}}{2}\right)}\, da\, db\ .$$

Let $\theta = b - \frac{\widetilde{R}}{2}a \qquad d\theta = db$

$$\lambda = a\sqrt{\frac{4 - \widetilde{R}^2}{4}} \qquad da = \frac{2}{\sqrt{4 - \widetilde{R}^2}}\, d\lambda$$

to obtain

$$E(pq) = \int_{-\infty}^{0}\int_{-\infty}^{-\frac{\widetilde{R}}{\sqrt{4 - \widetilde{R}^2}}} \frac{1}{\sqrt{2\pi}}\; e^{-\frac{\lambda^2+\theta^2}{2}}\, d\lambda\, d\theta$$

$$= \frac{1}{4} + \frac{1}{2\pi}\,\mathrm{Arc\ tan}\,\frac{\widetilde{R}}{\sqrt{4 - \widetilde{R}^2}}\ .$$

Then, using equation (5),

$$R = \frac{6}{\pi}\,\mathrm{Arc\ tan}\,\frac{\widetilde{R}}{\sqrt{4 - \widetilde{R}^2}}$$